



Universidade Federal de Mato Grosso

AULA 05 – Estatística

Prof. Lucas Bianchi

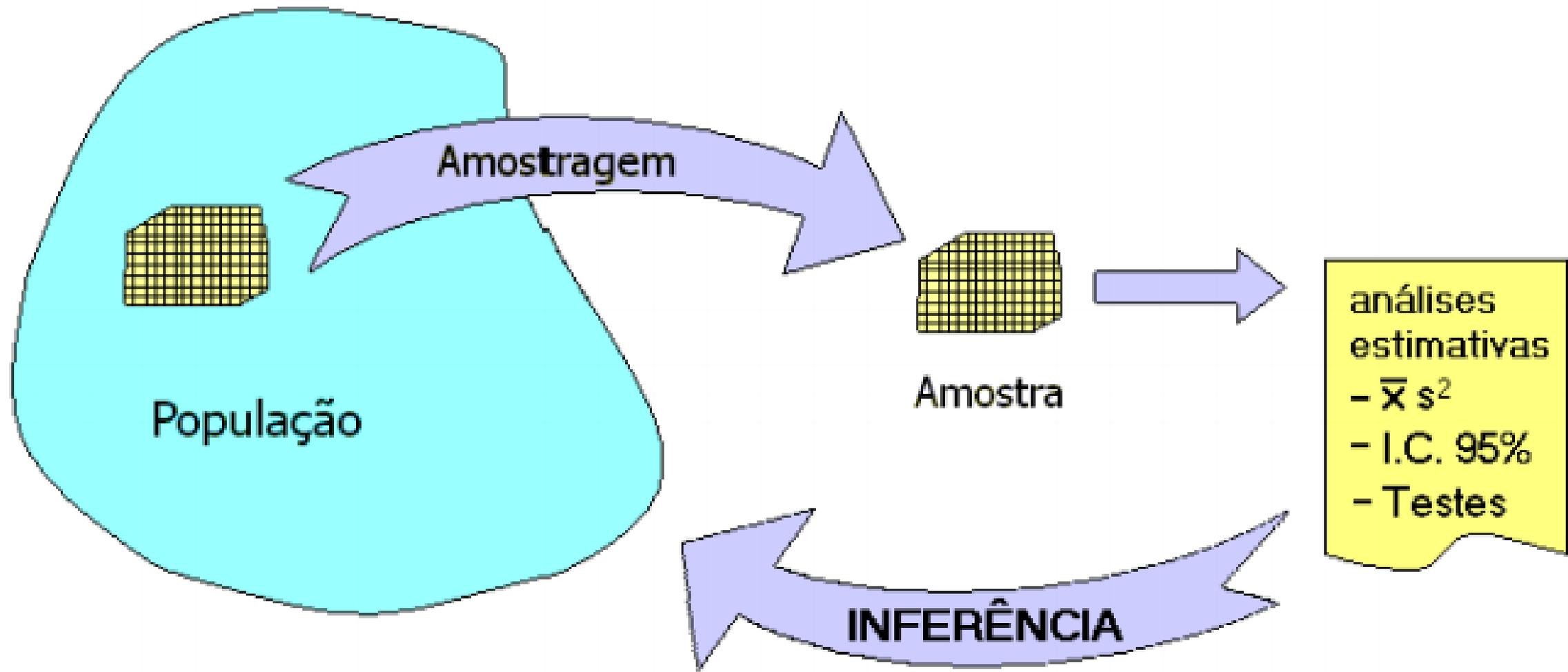
Cuiabá, 02 de setembro de 2016

Na aula de hoje, estudaremos:

- ✓ Amostragem não-probabilística
- ✓ Amostragem probabilística

Na realização de qualquer estudo quase nunca é possível examinar todos os elementos da população de interesse. Temos usualmente de trabalhar com uma amostra da população.

É errôneo pensar que, caso tivéssemos acesso a todos os elementos da população, seríamos mais precisos. Os erros de coleta e manuseio de um grande número de dados são maiores do que as imprecisões a que estamos sujeitos quando generalizamos, via inferência, as conclusões de uma amostra bem selecionada.



O que é amostragem?

É o processo de definição da amostra. É a parte inicial de qualquer estudo estatístico. Consiste na escolha criteriosa dos elementos a serem submetidos ao estudo.

Pode ser feito de maneira aleatória ou não. Quando bem utilizado é um fator responsável pela determinação da representatividade da amostra.

A preocupação central é que a amostra seja **representativa**.

Assim que decidimos obter informações através de um levantamento amostral, temos imediatamente dois problemas:

- Definir cuidadosamente a população de interesse;
- Selecionar a característica que iremos pesquisar.

A amostragem probabilística também é chamada de amostragem aleatória ou ao acaso. Este tipo de amostragem é submetida a tratamento estatístico que permite compensar erros amostrais.

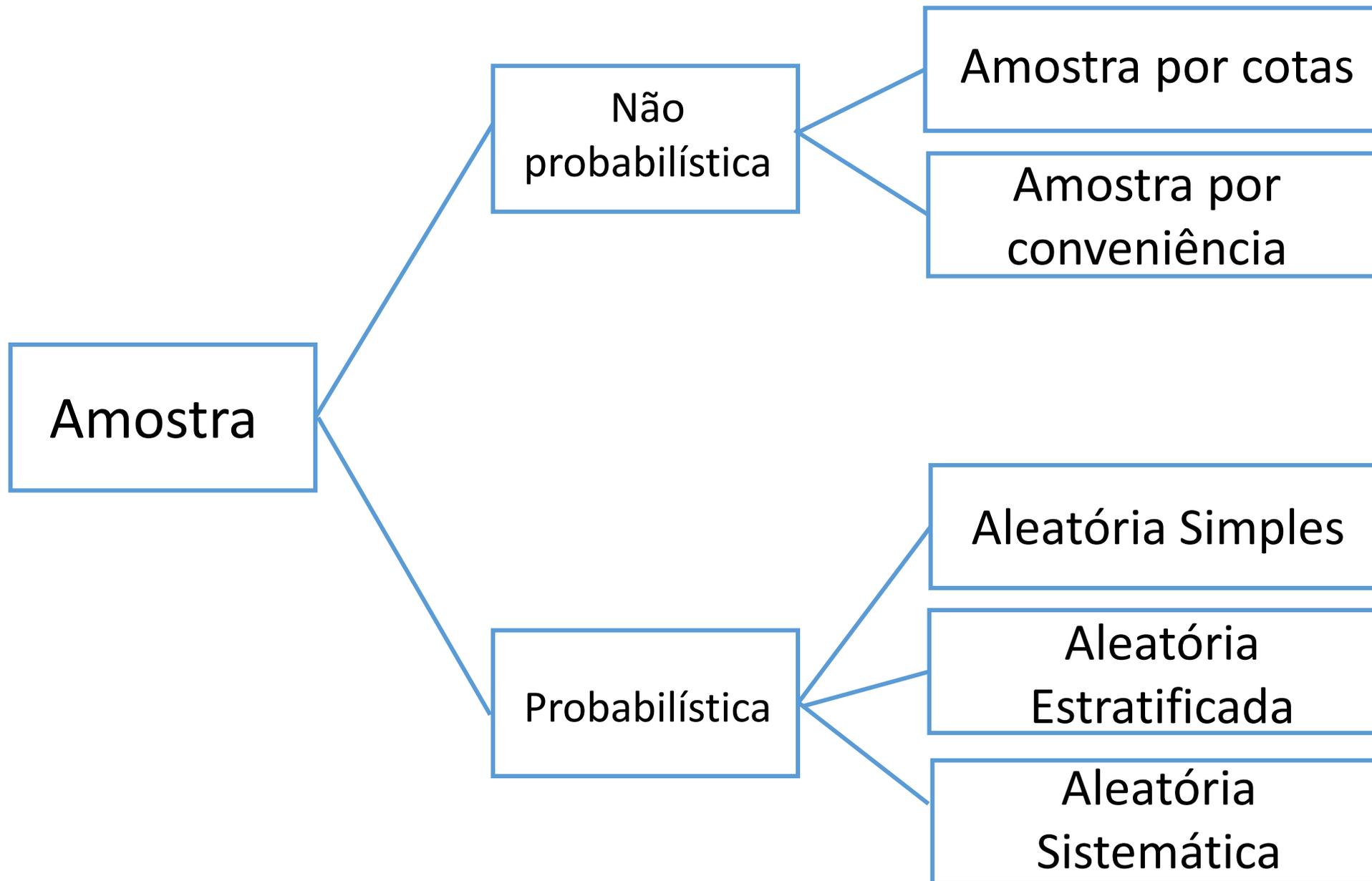
Hoje, dificilmente se aceita uma amostragem não-probabilística, exceto nos casos em que a amostragem probabilística não pode ser feita

Características importantes sobre a amostragem

- Usa-se a proporção de pessoas em uma amostra, portadoras de determinada característica, para estimar a proporção, na população das que tem essa característica.
- O melhor método de escolha de uma amostra é a escolha aleatória, isto é, que toda amostra possível tenha a mesma chance de ser escolhida.

Características importantes sobre a amostragem

- Antes de se proceder a observação de uma determinada população surge a questão se a amostragem será com ou sem reposição. Se o tamanho da amostra é insignificante em relação à população o impacto da reposição será desprezível, porém, se a amostra for grande então a reposição ou não pode causar um impacto significativo no resultado da probabilidade.
- Como as características das populações estatísticas variam, às vezes, é necessário se adequar esta população estatística para submetê-la a um critério de seleção possível, sem, contudo, perder seu caráter aleatório.



Amostragem não-probabilística

São amostragens em que há uma escolha deliberada dos elementos da amostra. Depende dos critérios e julgamento do pesquisador, ou seja, aquela em que a seleção dos elementos da população para compor a amostra depende ao menos em parte do julgamento do pesquisador ou do entrevistador no campo. Não há nenhuma chance conhecida de que um elemento qualquer da população venha a fazer parte da amostra.

Amostragem probabilística

As técnicas de amostragem probabilística, ou aleatória, ou ao acaso, são aquelas em que a seleção é aleatória de tal forma que cada elemento da população tem a mesma probabilidade conhecida e diferente de zero de ser selecionada para compor a amostra.

Principais Métodos de Amostragem

Como as características das populações variam, às vezes, é necessário submetê-la a um critério de seleção possível, sem, contudo, perder seu caráter aleatório.

- Amostragem Aleatória Simples;
- Amostragem Aleatória Sistemática;
- Amostragem Estratificada.

Amostragem Aleatória Simples (A.A.S)

A amostragem aleatória simples é o tipo de amostragem probabilística mais utilizada. Dá exatidão e eficácia à amostragem, além de ser o procedimento mais fácil de ser aplicado.

Neste tipo de amostra a premissa é de que cada componente da população estudada tem a mesma chance de ser escolhido para compor a amostra e a técnica que garante esta igual probabilidade é a seleção aleatória de indivíduos, por exemplo, através de sorteio.

Procedimento: Na prática a amostragem aleatória simples pode ser realizada numerando-se a população de 1 a N e sorteando-se, a seguir, por meio de um dispositivo aleatório qualquer, k números dessa sequência, os quais corresponderão aos elementos pertencentes á amostra.

Procedimento de sorteio:

1. Um indivíduo é selecionado ao acaso dentre os N possíveis;
2. O segundo indivíduo é selecionado ao acaso dentre os $(N - 1)$ restantes assim por diante, até que todos os n indivíduos sejam sorteados

Esse procedimento tem a característica de ser “sem reposição”, o que significa que: cada indivíduo aparece uma única vez na amostra.

Procedimentos “com reposição”, quando o indivíduo pode aparecer mais de uma vez na amostra, não serão abordados por serem poucos comuns na prática.

Obs: Quando o tamanho da população for muito grande, os dois procedimentos de sorteio (sem e com reposição) são equivalentes.

Como realizar o sorteio?

1. Geração números aleatórios, pelo computador;
2. Tabela de números aleatórios;
3. Globos com bolinhas numeradas;
4. Ou qualquer outra forma aleatória de escolha que preserve a propriedade de que cada unidade amostral tenha a mesma chance de ser selecionada.

Amostragem Aleatória Estratificada (A.A.E)

Quando a população é muito heterogênea, ou seja, quando as características observadas variam muito de um indivíduo para outro, é aconselhável subdividir a população em estratos homogêneos.

A população é dividida em k estratos sendo que, uma A.A.S. é aplicada em cada um dos deles.

Exemplo 1. Dada a população de 50.000 operários da indústria, selecionar uma amostra proporcional estratificada de 5% de operários para estimar seu salário médio. Usando a variável critério "cargo" para estratificar essa população, e considerando amostras de 5% de cada estrato obtido, chegamos a seguinte tabela:

Tabela 1: Amostragem Aleatória estratificada da indústria em estudo.

Cargo	População	5%	Amostra
Chefe de seção	5.000	$0,05*(5000)=250$	250
Operário Especializado	15.000	$0,05*(15000) =750$	750
Operário não Especializado	30.000	$0,05*(30000)=1500$	1.500
Total	50.000	$0,05*(50000)=2500$	2.500

Amostragem Aleatória Sistemática

Consiste em selecionar as unidades elementares da população em intervalos pré-fixados. Para funcionar, a técnica requer a listagem prévia da população, que deverá ser homogênea e uma atenção especial na periodicidade dos dados.

São exemplos os prontuários médicos de um hospital, os prédios de uma rua, uma linha de produção, os nomes em uma lista telefônica, etc.

Procedimento: Consideremos uma população, com elementos ordenados, de tamanho N e dela tiramos uma amostra de tamanho n da seguinte maneira:

1. Definimos o fator de sistematização (FS), dado por $FS = N/n$.
2. Sorteamos um número entre 1 e FS. Esse número é simbolizado por m , que será o primeiro elemento da amostra.
3. O segundo elemento da amostra é o de número $FS + m$.
4. O terceiro elemento da amostra é o de número $2FS + m$.
5. O k -ésimo elemento da amostra é o número $(k - 1) FS + m$

Exemplo 2. Uma rua contém 1000 prédios, dos quais desejamos obter uma amostra sistemática formada por 100 deles.

$$FS = 1000/100 = 10$$

m será um número entre 1 e 10. Vamos supor que $m = 7$. Então temos:

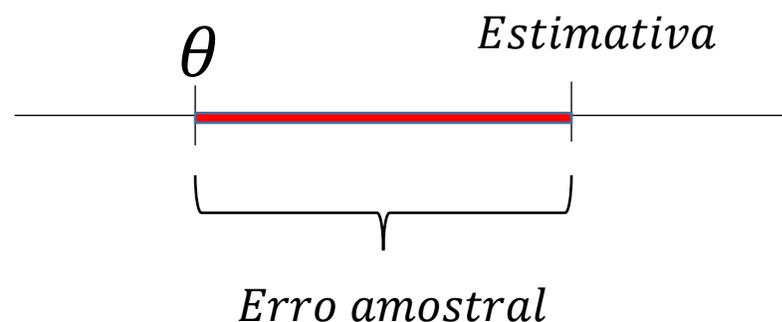
1º elemento da amostra = $(1 - 1)10 + 7 = 7 > 7^\circ$ elemento da população;

2º elemento da amostra = $(2 - 1)10 + 7 = 17 > 17^\circ$ elemento da população;

100º elemento da amostra = $(100 - 1)10 + 7 = 997 \Rightarrow 997^\circ$ elemento da população.

Erro Amostral

O erro amostral é definido como sendo a diferença entre a estimativa obtida para um parâmetro e o seu verdadeiro valor. É decorrente da variabilidade natural das unidades amostrais (é aleatório).



Alguns erros em amostragem:

1. População acessível diferente da população alvo;
2. Erros de mensuração.

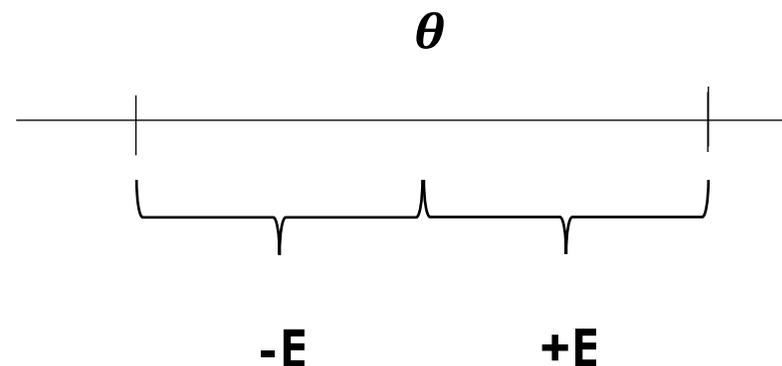
Tamanho da amostra

A determinação do tamanho da amostra é, talvez, o grande dilema dos pesquisadores, pois deve levar em conta um erro tolerável e a probabilidade de se cometer tal erro.

O erro tolerável é uma margem de erro das estimativas em relação ao parâmetro θ , para mais ou para menos, o qual o pesquisador está disposto a aceitar.

O tamanho da amostra é determinado tal que a probabilidade de que a estimativa do parâmetro θ esteja dentro da margem de erro seja alta, por exemplo, de 95%

$$P(\text{estimativa de } \theta \text{ estar dentro da margem de erro})=0.95$$



Ao se obter uma amostra qualquer de tamanho n , calcula-se a média aritmética. Provavelmente, se uma nova amostra aleatória for realizada, a média aritmética obtida será diferente daquela da primeira amostra. A variabilidade das médias é estimada pelo seu erro padrão.

Assim, o erro padrão avalia a precisão do cálculo da média populacional.

O erro padrão é dado pela fórmula:

$$E = \frac{s}{\sqrt{n}}$$

Onde:

$E \rightarrow$ é o erro padrão (margem de erro)

$s \rightarrow$ é o desvio padrão

$n \rightarrow$ é o tamanho da amostra

Observação: quanto melhor a precisão no cálculo da média populacional, menor será o erro padrão.

Exemplo 3. Numa população obteve-se desvio padrão de 2,64 com uma amostra aleatória de 60 elementos. Qual o provável erro padrão?

Solução:

$$E = \frac{s}{\sqrt{n}} = \frac{2,64}{\sqrt{60}} = 0,3408$$

Isso indica que a média pode variar 0,3408 para mais ou para menos.

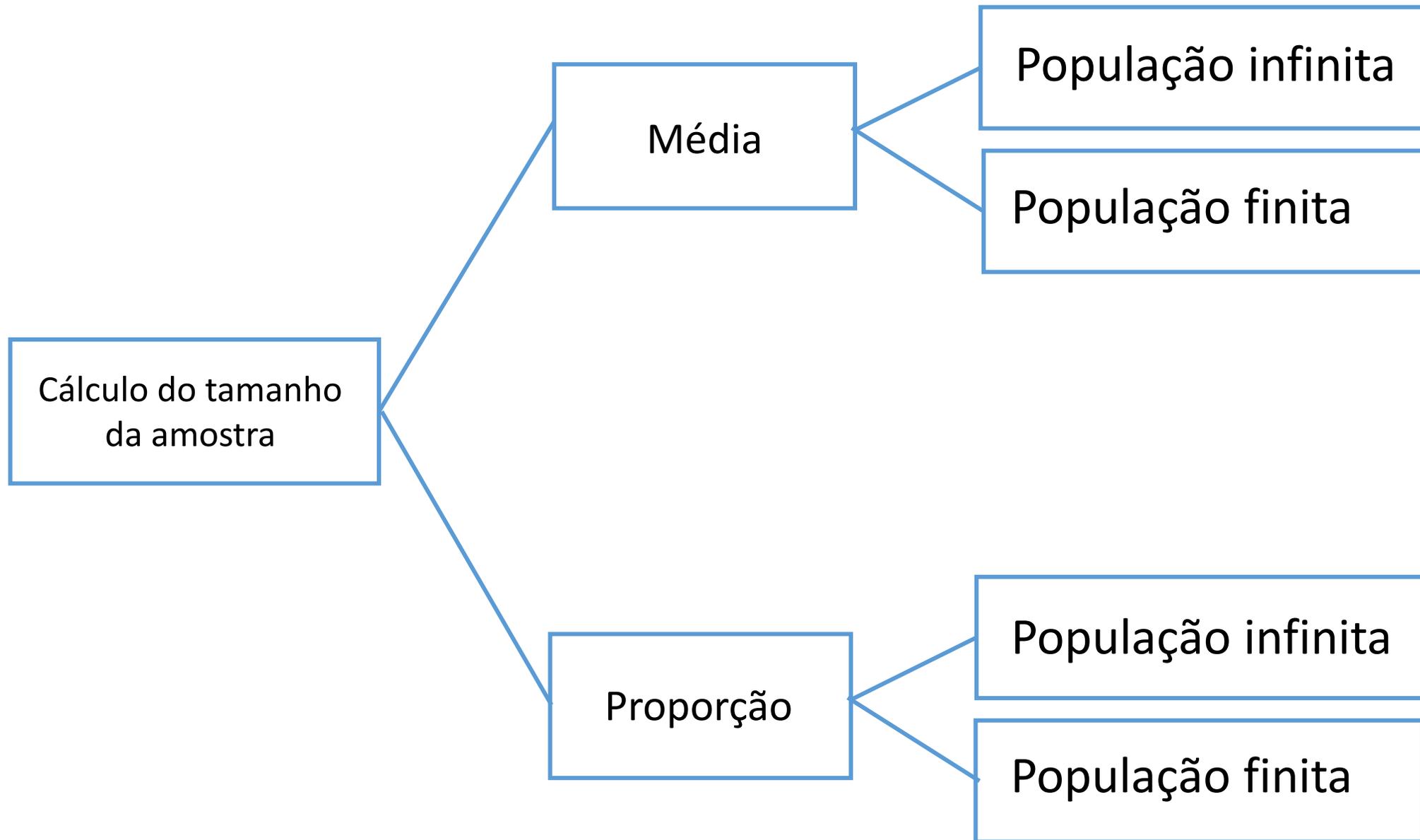
Exemplo 4. Numa população obteve-se desvio padrão de 1,32 com uma amostra aleatória de 121 elementos. Sabendo que para essa mesma amostra obteve-se uma média de 6,25, determine o valor mais provável para a média dos dados.

Solução: Para determinarmos o valor mais provável da média dos dados devemos calcular o erro padrão da estimativa. Assim, teremos:

$$E = \frac{s}{\sqrt{n}} = \frac{1,32}{\sqrt{121}} = 0,12$$

Finalizando, o valor mais provável para a média dos dados obtidos pode ser representado por:

$$\bar{x} = 6,25 \pm 0,12$$



Tamanho da amostra - Média Populacional

População Infinita: A fórmula para cálculo do tamanho da amostra para uma estimativa confiável da média populacional (μ) é dada por:

$$n = \left(\frac{Z_{\alpha/2} \sigma}{E} \right)^2$$

População Finita: A fórmula para cálculo do tamanho da amostra para uma estimativa confiável da MÉDIA POPULACIONAL (μ) é dada por:

$$n = \frac{N \cdot \sigma^2 \cdot (z_{\alpha/2})^2}{(N - 1) \cdot E^2 + \sigma^2 \cdot (z_{\alpha/2})^2}$$

A fórmula para cálculo do tamanho da amostra para uma estimativa confiável da MÉDIA POPULACIONAL (μ) é dada por:

$$n = \left(\frac{Z_{\alpha/2} \sigma}{E} \right)^2$$

n = Número de indivíduos na amostra

$Z_{\alpha/2}$ = Valor crítico que corresponde ao grau de confiança desejado.

σ = Desvio-padrão populacional da variável estudada

E = Margem de erro

Exemplo: Um economista deseja estimar a renda média para o primeiro ano de trabalho de um bacharel em direito. Quantos valores de renda devem ser tomados, se o economista deseja ter 95% de confiança em que a média amostral esteja a menos de R\$500,00 da verdadeira média populacional? Suponha que saibamos, por um estudo prévio, que para tais rendas, $\sigma = \text{R}\$6250,00$.

$$Z_{\alpha/2} = 95\% = 1,96$$

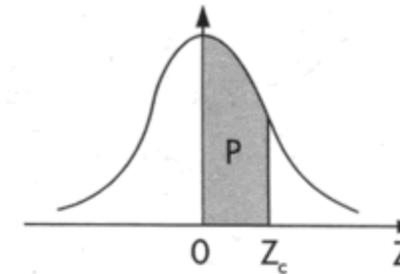
$$\sigma = \text{R\$ } 6250,00$$

$$E = \text{R\$ } 500,00$$

$$n = \left(\frac{Z_{\alpha/2} \sigma}{E} \right)^2 = \left(\frac{1,96 * 6.250}{500} \right)^2 = 600,25 = 601$$

Devemos, portanto, obter uma amostra de ao menos 601 rendas do primeiro ano dos bacharéis, selecionadas aleatoriamente.

Tabela III – Distribuição Normal Padrão
 $Z \sim N(0, 1)$
 Corpo da tabela dá a probabilidade p , tal que $p = P(0 < Z < Z_c)$



parte inteira e primeira decimal de Z_c	Segunda decimal de Z_c										parte inteira e primeira decimal de Z_c
	0	1	2	3	4	5	6	7	8	9	
	p = 0										
0,0	00000	00399	00798	01197	01595	01994	02392	02790	03188	03586	0,0
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535	0,1
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409	0,2
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173	0,3
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793	0,4
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240	0,5
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490	0,6
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524	0,7
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327	0,8
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891	0,9
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214	1,0
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298	1,1
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147	1,2
1,3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774	1,3
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189	1,4
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408	1,5

	0	1	2	3	4	5	6	7	8	9	
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449	1,6
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327	1,7
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062	1,8
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670	1,9
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169	2,0
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574	2,1
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899	2,2
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158	2,3
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361	2,4
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520	2,5
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643	2,6
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736	2,7
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807	2,8
2,9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861	2,9
3,0	49865	49869	49874	49878	49882	49886	49889	49893	49897	49900	3,0
3,1	49903	49906	49910	49913	49916	49918	49921	49924	49926	49929	3,1
3,2	49931	49934	49936	49938	49940	49942	49944	49946	49948	49950	3,2
3,3	49952	49953	49955	49957	49958	49960	49961	49962	49964	49965	3,3
3,4	49966	49968	49969	49970	49971	49972	49973	49974	49975	49976	3,4
3,5	49977	49978	49978	49979	49980	49981	49981	49982	49983	49983	3,5
3,6	49984	49985	49985	49986	49986	49987	49987	49988	49988	49989	3,6
3,7	49989	49990	49990	49990	49991	49991	49992	49992	49992	49992	3,7
3,8	49993	49993	49993	49994	49994	49994	49994	49995	49995	49995	3,8
3,9	49995	49995	49996	49996	49996	49996	49996	49996	49997	49997	3,9
4,0	49997	49997	49997	49997	49997	49997	49998	49998	49998	49998	4,0
4,5	49999	50000	50000	50000	50000	50000	50000	50000	50000	50000	4,5

Tabela dos valores críticos associados ao grau de confiança da amostra

Grau de confiança	α	Valor Critico
90%	0,1	1,64
95%	0,05	1,96
99%	0,01	2,57

Exemplo 5. Utilizando as informações do exemplo 4, uma população obteve-se desvio padrão de 1,32 com uma amostra aleatória de 121 elementos. Com 95% de confiança, qual é o tamanho da amostragem? E com 90% e 99% de confiança?

$$E = \frac{s}{\sqrt{n}} = \frac{1,32}{\sqrt{121}} = 0,12$$

$$n = \left(\frac{Z_{\alpha/2} \sigma}{E} \right)^2 = \left(\frac{1,96 * 1,32}{0,12} \right)^2 = 464,8336 \cong 465$$

A fórmula para cálculo do tamanho da amostra para uma estimativa confiável da MÉDIA POPULACIONAL (μ) é dada por:

$$n = \frac{N \cdot \sigma^2 \cdot (z_{\alpha/2})^2}{(N - 1) \cdot E^2 + \sigma^2 \cdot (z_{\alpha/2})^2}$$

onde:

n = Número de indivíduos na amostra

$Z_{\alpha/2}$ = Valor crítico que corresponde ao grau de confiança desejado.

σ^2 = Variância populacional da variável estudada

E = Margem de erro ou ERRO MÁXIMO DE ESTIMATIVA.

N = População Total

Exemplo 6. O município de Arapiraca das Dornas deseja fazer uma pesquisa do peso de papéis descartados mensalmente pelas residências da cidade, para efeito de planejamento da coleta de lixo. O peso médio de papel descartado, por uma amostra teste de 31 residências, foi de 9,4281 kg. e o desvio padrão dessa amostra foi de 4,1681 kg. Como a cidade possui 2.637 residências, deseja-se calcular o tamanho da amostra a ser coletada, com grau de confiança de 95 %.

Conhecemos o desvio padrão amostral, s , que, para 62 indivíduos, substitui o desvio padrão populacional. Trata-se de uma amostra finita, na qual:

$$N = 2.637 \quad \bar{x} = 9,4281$$

$$s = 4,1681 \quad E = \frac{s}{\sqrt{n}} = 0,7471$$

$$Z_{\alpha/2} = 95\% = 1,96$$

Conhecemos essas informações, basta aplicar a formula.

$$n = \frac{N \cdot \sigma^2 \cdot (z_{\alpha/2})^2}{(N - 1) \cdot E^2 + \sigma^2 \cdot (z_{\alpha/2})^2}$$

$$n = \frac{2637 \cdot (4,1681)^2 \cdot (1,96)^2}{(2637 - 1) \cdot (0,7471)^2 + (4,1681)^2 \cdot (1,96)^2}$$

$$n \cong 114,4271 \cong 115$$

Exercício 1. E se essa cidade tivesse 5000 residências e desvio padrão de 12,5kg de papel, qual seria o tamanho da Amostra?

$$n = \frac{N \cdot \sigma^2 \cdot (z_{\alpha/2})^2}{(N - 1) \cdot E^2 + \sigma^2 \cdot (z_{\alpha/2})^2}$$

$$n = \frac{5000 \cdot (12,5)^2 \cdot (1,96)^2}{(5000 - 1) \cdot (0,7471)^2 + (12,5)^2 \cdot (1,96)^2}$$

$$n \cong 885,1981 \cong \mathbf{886}$$

Tamanho da amostra - Proporção Populacional

População Infinita: A fórmula para cálculo do tamanho da amostra para uma estimativa confiável da PROPORÇÃO POPULACIONAL (p) é dada por:

$$n = \frac{Z_{\alpha/2}^2 \cdot \hat{p} \cdot \hat{q}}{E^2}$$

População Finita: A fórmula para cálculo do tamanho da amostra para uma estimativa confiável da PROPORÇÃO POPULACIONAL (p) é dada por:

$$n = \frac{N \cdot \hat{p} \cdot \hat{q} \cdot (z_{\alpha/2})^2}{\hat{p} \cdot \hat{q} \cdot (z_{\alpha/2})^2 + (N - 1) \cdot E^2}$$

Onde:

\hat{p} = Proporção populacional de indivíduos que pertence a categoria que estamos interessados em estudar.

\hat{q} = Proporção populacional de indivíduos que NÃO pertence à categoria que estamos interessados em estudar $q = (1 - p)$.

Exemplo 7. Um assistente social deseja saber o tamanho da amostra (n) necessário para determinar a proporção da população atendida por uma Unidade de Saúde, que pertence ao município de Jaciara. Não foi feito um levantamento prévio da proporção amostral e, portanto, seu valor é desconhecido. Ela quer ter 90% de confiança que sua o erro máximo de estimativa (E) seja de $\pm 5\%$ (ou 0,05). Quantas pessoas necessitam ser entrevistadas?

Importante: Quando não há informação sobre a população, geralmente utilizamos $p = 50\%$

$$n = \frac{Z_{\alpha/2}^2 \cdot p \cdot q}{E^2} = \frac{1,64^2 \cdot 0,5 \cdot 0,5}{0,05^2} = 268,96 = 269$$

Devemos, portanto, obter uma amostra de 269 pessoas para determinar a proporção da população atendida na Unidade de Saúde, que se origina do município de Jaciara.

Exemplo 8. Utilizando o mesmo exemplo, vamos considerar que a população tenha tamanho 5000, qual é o tamanho da amostra?

$$n = \frac{N \cdot \hat{p} \cdot \hat{q} \cdot (z_{\alpha/2})^2}{\hat{p} \cdot \hat{q} \cdot (z_{\alpha/2})^2 + (N - 1) \cdot E^2}$$

$$n = \frac{5000 \cdot 0,5 \cdot 0,5 \cdot (1,64)^2}{0,5 \cdot 0,5 \cdot (1,64)^2 + (5000 - 1) \cdot 0,05^2} = 255,2790 = 256$$

Exercício 2. Considere a quantidade de alunos presentes na aula de hoje. Calcule o tamanho da amostra necessária para que essa amostragem seja representativa com 90%, 95% e 99% de confiança e erro de 5%.

Exercício 3. Se aplicarmos amostragem estratificada por sexo com base no cálculo do exercício 2 para 95% de confiança, quantas pessoas precisariam ser amostradas por estrato.

Exercício 4. Se consideramos a utilização da amostragem sistemática, qual seria o fator de sistematização?